REAGENTS AND METHODS FOR DIVERSIFICATION OF DNA

Field of the invention

The present invention relates to the field of molecular biology and a method and reagents for the diversification of DNA sequences. The present invention allows the generation of variant DNA sequences, which can provide variants of sequences that regulate gene expression, code for proteins, control the export of gene products from cells and the localization of gene products within cells. In particular, the present invention provides a method and reagents for efficiently generating new variant sequences in heterologous DNA (DNA foreign to the cell) by placing two sequences, that differ at multiple sites, into the same cell at a location that permits the expression of the new variant sequence after diversification and which provides for the generation of new variants as a result of the exchange of parts of the two sequences within the living cell.

Background of the Invention

Diversification of DNA molecules provides a way to generate new proteins with properties not found in nature. This can be achieved *in vitro* (by manipulation outside living cells) by a variety of methods such as the insertion of specific novel and random sequence oligonucleotides (Gold, L *et al* 1997 Proceedings of the National Academy of Science, 94:59-97) in selected regions of genes, or the cleavage of variant DNA molecules (two sequences that have similar functions but differ in one or more sites) and reassembly of the fragments in new combinations (Stemmer, WPC 1994 Proceedings of the National Academy of Science, 91: 10747-10751), or by amplification of DNA by the polymerase chain reaction under conditions where the polymerase is error prone (Leung, DW *et al* 1989 Technique 1: 11-15). Screening for desired properties of any protein coded by the resultant novel nucleotide sequence requires transcription and translation of the sequence to yield the corresponding peptide and the appropriate post-translational modification. In these cases, this is usually achieved by introducing each new construct into a cell by

transfection or electroporation (hereafter both processes will be covered by the term transfection) to form a transformed cell, a complex and time consuming procedure since it is necessary to check each construct to ensure it is correctly inserted and is complete. This difficulty is compounded where the new construct codes for one component of a multimeric protein since the procedure must be done twice for each new combination. That problem can be reduced by the use of fungal heterokaryons to lower the number of transfections required to one per component of a combinatorial array (US Patent Serial No. 5,643,745, to Stuart, issued July 1 1997). However, the number of transfections required is still large and each must be checked to ensure the DNA insert is correctly placed and complete.

Genetic recombination

Genetic recombination in eukaryotes, higher organisms that have a true nucleus, occurs during the prophase of the reduction division that converts a diploid cell having two complete sets of homologous chromosomes to a tetrad (or sometimes an octad) of haploid cells each with one complete set of chromosomes. Two manifestations of recombination events are recognized: crossing over in which genes located at different sites (loci) on the same chromosome are recombined by reciprocal exchange of chromosome sections between a pair of homologous chromosomes and gene conversion in which the number of copies of a pair of allelic genes, ie genes that occupy the same locus on homologous chromosomes, is unequal in the tetrad or octad. Instead of a two:two segregation of the parental alleles, the tetrad comprises three haploid cells carrying one of the parental versions of the gene and one carrying the other parental version of the gene. Crossing over was first discovered in the fruit-fly Drosophila (Morgan Proc. Soc. Exp. Biol. Med. **8:**17 1910) and gene conversion in the fungus Neurospora (MB.Mitchell Proc. Natl. Acad. Sci. USA **41:** 216-220 1955). There is now evidence that both crossing over and gene conversion occur universally in species that reproduce sexually and that a process having similar outcomes occurs in bacteria and their viruses and plasmids.

Genetic recombination in eukaryotes occurs in diploid cells (cells that contain two complete sets of homologous chromosomes) that are undergoing meiosis. Prior to the

division, each of the two chromosome sets is replicated, generating two pairs of identical sister chromatids. The process of genetic recombination involves the establishment of joints between two homologous but not necessarily identical DNA sequences, one located on one chromatid of one sister pair and the other in a homologous chromatid which is a member of the other sister pair. The joints establish regions in one or both chromatids where one strand of the DNA duplex has the sequence of one homologue and the second strand has the sequence of the other homologue. Where the DNA sequences of the homologues differ, bases will be in mismatched pairs, that is pairs which are not A:T or G:C (A=deoxyadenine, T=deoxythymine, C=deoxycytidine and G=deoxyguanine). Enzymatic machinery corrects mismatched base pairs and the joints between the molecules are resolved, separating the two chromatids once more. For each site of mismatch, in half of the cases, the base pair present in one chromatid is now replaced by the base pair originally present in the other homologous chromatid. This accounts for gene conversion. In some cases the joints between molecules are resolved such that there is a reciprocal exchange of the regions each side of the joint. This process is called crossing over and also leads to novel combinations of DNA sequence information by which the parental homologues differed. Each chromatid is incorporated into one of the haploid cells (cells having only one set of chromosomes) that arise from meiosis, becoming a member of the complete set of chromosomes present in each cell.

The molecular processes of crossing over and gene conversion are yet to be fully understood. In the most widely accepted model for the molecular events of recombination (Figure 3) (H Sun *et al* Cell 64: 1155-1161, 1991) it is supposed that one of the two homologous chromatids suffers a break in both strands of the DNA molecule and that the strands that end with a 5' phosphate are resected, leaving a single strand tail of several hundred bases that ends with a 3' hydroxyl group. It is proposed that the single strand tail pairs with the complementary strand of the unbroken chromosome to initiate the joint. The joint is thought to be completed by DNA synthesis from the 3' ends to provide a replacement strand for the DNA lost in the initial resection followed by rejoining of the breaks. This will form a double junction between the molecules in the manner shown in Figure 3. Each junction is free to move. This leads to strand exchange between the two

DNA molecules forming heteroduplex DNA. It is supposed that recombination is completed by scission of the junctions and correction of mispaired bases. Scission of the junctions can occur by breaks in either the "inner" or "outer" strands with equal probability (Figure 3). Due to the limitations of a two dimensional representation of the junctions, the expectation of an equal frequency of these two modes of scission is not self evident. However in reality, the two pairs of complementary strands, both the inner and outer pair, are identically juxtaposed. If the resolution of both junctions occurs in the inner strands or alternatively in the outer strands, only gene conversion can occur. If the resolution of one junction is by scission of the inner strands and the other junction by scission of the outer strands, the flanking regions are reciprocally exchanged and there is both a crossover event and also the possibility of gene conversion.

There is direct evidence that recombination is initiated by two strand breaks in the yeast *Saccharomyces cereviseae* (A Schwacha and N Kleckner, Cell **83:** 1-20 1995). However, the exact series of events by which these are processed to complete a recombination event is not clear. Indeed, Bowring and Catcheside working with the fungus *Neurospora crassa* (Genetics **143:** 129-136 1996) have shown that most of the crossing over events previously thought to be associated with gene conversion are several hundred kilobases away, too far to be directly associated, suggesting that gene conversion and crossing over can be catalyzed by different recombination pathways.

Biological processes including recombination are error prone. MK Watters and DR Stadler (Genetics 139, 137-145 1995) examined the spectrum of spontaneous mutations (changes in the sequence of DNA bases in a gene, from that present in wild-type cells, that render it defective) in the *mtr* gene of *Neurospora crassa*. Watters and Stadler found that the spectrum of mutations which occur during the sexual phase that includes meiosis and recombination is distinct from those that occur during asexual reproduction by normal vegetative growth. Error prone recombination is a source of sequence diversification *in vivo* additional to that obtainable by the generation of new combinations of multiple sequence differences that distinguish homologous DNA sequences.

Genetic recombination in eukaryotes occurs in diploid cells that contain two

complete sets of chromosomes and thus two complete sets of genes. The diploid state is established by the fusion of two haploid cells, usually of different parentage. This can be achieved by the fusion of gametes, as in the fusion of eggs and sperm in humans and other animals or of pollen cells with ovules in plants, or by fusion of two strains in the fungi where ability to fuse is usually controlled by mating type genes that ensure those strains that fuse are of different mating type and thus not genetically identical. In plants and animals, the fusion of haploid gametes establishes a clone of diploid cells which normally develops into an individual adult member of the species where genetic recombination occurs in specialist diploid cells in those parts of adults that give rise either to eggs or sperm. In the fungi, fusion of haploid strains usually gives rise to a dikaryon (a cell having haploid nuclei of two types, each with the genetic composition of one of the two strains, in a common cytoplasm). The dikaryon can form the main phase of the life cycle, as in the macrofungi ("mushrooms" and "toadstools"), or can be transient and give rise to diploid cells, immediately or after a limited number of mitotic cell divisions, that then undergo meiosis.

Genetic recombination in eukaryotes occurs during meiosis, the reduction division in which a diploid nucleus gives rise to four haploid nuclei each having only one set of chromosomes. During this process, the genetic information in the two sets of chromosomes present in the nucleus of the diploid cell is recombined. New gene combinations can be generated by reassortment of chromosomes between the sets present in the two haploid cells that contributed to the diploid cell undergoing meiosis and also by crossing over and gene conversion which generate new combinations of the sequence information present in pairs of homologous chromosomes. In prokaryotes, genetic recombination can occur between DNA sequences present in the chromosome and those carried by plasmids such as the fertility factor F of *Eschericia coli* or bacteriophage such as phage λ (*lambda*) and between two phage molecules, two plasmids or any combination thereof.

The methods and compositions to cross together two genetically distinct individuals or strains of a living organism in order to obtain individuals with new gene

combinations by reassortment, crossing over and gene conversion varies from species to species and for most species is within the common art of the biological sciences. Some species are better characterized genetically than are others, as a result of their being of particular economic importance, particular ecological or aesthetic importance or are species that are particularly favorable for research into the fundamental processes of biology. The best characterized species include the bacterium *Escherichia coli*, the plants *Arabidopsis thaliana* and *Oryza sativa*, the insect *Drosophila melanogaster*, the mammal *Mus musculus*, the nematode *Cenorhabditis elegans*, the slime mould *Dictyostelium discoideum* and the fungi *Saccharomyces cereviseae*, *Aspergillus nidulans* and *Neurospora crassa*. In each case there are compendia of standard methods for their growth and for conducting crosses. For example for *Neurospora crassa*, these include DD Perkins *et al* (Microbiol. Rev. 46:426-570 1982) RH Davis and FJ deSerres (Methods in Enzymol. 17A: 79-143 1970). The following details of methods and compositions for genetic recombination in the fungi are given as examples and are not intended to limit the application of the invention for *in vivo* diversification of DNA sequences to these species. Nevertheless, bacteria are in general not suitable for the purpose of diversifying and expressing eukaryote sequences due to the lack of the correct processing pathways for proper gene expression and modification of any protein product. Amongst the eukaryotes, only in the fungi has understanding of the relevant molecular processes reached the level required for practical application of the present invention.

Recombination Hotspots

Crossing over and gene conversion during meiosis do not occur at random positions within chromosomes. Recombination is particularly frequent in regions called recombination hotspots. Recombination hotspots are also called recombinators. Recombination hotspots typically occur at several locations on a chromosome, frequently, but not always being in the regulatory region 5' of the coding sequence of a gene. They have been directly demonstrated in several species including the yeasts, *Schizosaccharomyces pombe* at the *ade6* gene and *Saccharomyces cerevisiae* at the *arg4* and *his4* loci (M Lichten and ASH Goldman Ann. Rev. Genet. 29: 423-444 1995) and in

the filamentous fungi in the Ascomycete *N. crassa* at *cog* (DG Catcheside & T Angel Aust. J. Biol. Sci 27: 219-229 1974) and at the *am* and *his-3* loci and in the Basidiomycete *Schizophilum commune* (G Simchen and J Stamberg Heredity 24: 369-381 1969) at mating type loci . Recombination hotspots that have been studied include the *arg4* and *his4* hotspots in yeast. The ability of yeast recombinators to diversify heterologous DNA has not been demonstrated, and further, unlike *cog* and other recombinators in Neurospora, the yeast recombinators have not been shown to be regulated.

There is indirect evidence that recombination hotspots are widely distributed in higher eukaryotes including *Homo sapiens* (KF Lindahl Trends. Genet. 7: 273-276 1991) and plants including *Zea mays* (L Civardi *et al* Proc. Nat. Acad. Sci. USA 91: 8268-8272 1994). Recombination hotspots in bacteria include χ (*chi*) (RS Myers and FW Stahl, Ann. Rev. Genet. 28: 49-70 1995) which stimulates recombination between any pair of bacterial chromosomes, phages or plasmids and site specific recombinators such as *att* which stimulate insertion and excision of phage such as phage λ (*lambda*).

In the case of the filamentous fungi, it is known that at least some of the recombination hotspots are subject to regulatory genes that turn them off. The genetic systems that regulate hotspot activity are well known only in Neurospora where the genes *rec-1, rec-2* and *rec-3* each turn off a different subset of hotspots scattered in the Neurospora genome. *rec-1* blocks recombination at the *nit-2* and *his-1* loci. *rec-2* blocks recombination at the *his-3* locus and also in the chromosomal regions between the *his-3* and *ad-3*, *arg-3* and *sn* and *pyr-3* and *his-5* loci. *rec-3* blocks recombination at the *am* and *his-2* loci. Control of recombination by *rec* genes in *N. crassa* has been reviewed by DEA Catcheside (Genetical Research, 47: 157-165 1986).

There remains a need for an effective reagents for and methods employing the process of recombination and recombination hot spots to introduce sequence variation into, to diversify, heterologous DNA.

Summary of the Invention

The present invention relates to fungal cells, reagents, methods, and the like for

diversification of DNA. Preferably the DNA to be diversified is heterologous DNA introduced into a plasmid and/or a fungal cell. The fungal cell of the invention can be either a diploid or a haploid fungal cell having a recombinant genome. A haploid fungal cell of the invention includes heterologous DNA functionally coupled to a recombination hot spot or recombinator. The haploid cell can be converted to a diploid cell, and the heterologous DNA can undergo recombination in the diploid cell. A diploid fungal cell of the invention includes a first and second heterologous DNA, each of which is functionally coupled to a first and second recombination hot spot, respectively. The first and second heterologous DNA can recombine.

A plasmid of the invention is suitable for replication of heterologous DNA in a fungal cell. A preferred plasmid includes a truncated *N. crassa his-3* gene and a *N. crassa* recombination hot spot, both of which are functionally coupled to a heterologous DNA. A preferred plasmid also includes a multiple cloning site 3' to the *his-3* gene and a marker gene. Advantageously, the plasmid can transfect an *N. crassa* cell. The plasmid can be incorporated in a fungal cell.

Diversified DNA can be prepared by using one or more of a plasmid of the invention or a fungal cell of the invention in a method of the invention. The method of the invention provides for introducing heterologous DNA into a fungal cell for recombination, crossing over, and/or conversion. This results in diversification of the heterologous DNA. A preferred method includes the steps of constructing strains of a fungus including heterologous DNA to be diversified with each heterologous DNA coupled to a corresponding recombination hot spot. This preferred method includes mating the strains to form a dikaryon, establishing a diploid cell line from the dikaryon, and inducing meiosis. Advantageously, meiosis includes one or more of gene conversion, crossing over, errors in recombination. This results in diversifying the heterologous DNA. The invention also includes a kit for carrying out the method.

A strain of fungus, preferably a strain useful for producing diversified DNA, can be formed by another method of the invention. This method requires making a diploid fungal cell containing non-complementary alleles of a gene providing an auxotrophic mutation, which gene is functionally coupled to, adjacent to, or juxtaposed to the

heterologous DNA to be diversified. This diploid allows enrichment for cells containing diversified DNA. Preferably, a first fungal cell having a first allele of the gene providing an auxotrophic mutation is transfected with a plasmid including a first heterologous DNA and a second allele of the gene providing an auxotrophic mutation. Advantageously, each of the first and second alleles encode a defective gene but are complementary alleles. The presence of the first and second allele of a gene following transfection of the first fungal cell establishes a first heterokaryon, which is grown to provide a first homokaryon containing the second allele of the gene and the first heterologous DNA.

Preferably, a second fungal cell having a third allele of the gene providing an auxotrophic mutation is transfected with a plasmid including a second heterologous DNA and a fourth allele of the gene providing an auxotrophic mutation. Advantageously, each of the third and fourth alleles encode a defective gene but are complementary alleles. The presence of the third and fourth allele of the gene providing an auxotrophic mutation following transfection of the second fungal cell establishes a second heterokaryon, which is grown to provide a second homokaryon containing the fourth allele of a gene and the second heterologous DNA. In a preferred embodiment, the first allele and the third allele are the same allele. The desired fungal strain is established by crossing the first and second homokaryons.

Brief Description of the Drawings

Figure 1 illustrates methods for diversification of DNA sequences and testing for variants.

Figure 2 illustrates methods for diversification of DNA sequences coding subunits of heteropolymeric proteins and testing for variants.

Figure 3 illustrates a modified double strand break repair model for meiotic recombination.

Figure 4 illustrates the lifecycle of *Neurospora crassa*.

Figure 5 illustrates a map of the *his-3, cog, lpl* region of linkage group 1 of *Neurospora crassa*.

Figure 6 illustrates discontinuity in the parental origin of DNA sequences and progeny from crosses between pairs of *his-3* alleles.

Figure 7 illustrates the nucleotide sequence of the *his-3 cog^{L} lpl* region of linkage group 1 of the Lindegren wild type strain of *Neurospora crassa* (SEQ. ID NO:1).

Figure 8 shows the nucleotide sequence of the *his-3 cog^{E} lpl* region of linkage group 1 of the St Lawrence wild type strain of *Neurospora crassa* (SEQ. ID NO:2).

Figure 9 illustrates construction of the components of a sequence diversification cross.

Figure 10 illustrates construction of parent strains for crossing.

Figure 11 illustrates sequence polymorphism between F11089 and F6325 at am and PCR primer sets used in sequencing and genotype determinations.

Figure 12 illustrates conversion frequency of am markers. Data are for am^{6} convertants and were extrapolated to include progeny for which molecular markers were not scored. A′ is the conversion frequency of A excluding events associated with discontinuous tracts.

Figure 13 illustrates minimum conversion tract coverage. Data are for am^{6} convertants and were extrapolated to include progeny for which molecular markers were not scored (see hereinbelow). Position relative to the first base of the first am codon is shown in basepairs at the top. Peak is the position of the inferred peak of conversion events. Each estimate is a minimum since the spacing of markers does not allow for exact positioning of tract ends.

Figure 14 illustrates the constitution of B163 parents and prototrophic offspring. Data for sp, HP, HD, 24HP, 24HD1, and his-1 are from (Bowring and Catcheside, 1996). Positions of markers relative to the first base of the first am codon, in basepairs unless otherwise indicated, are shown above marker designations. All markers except sp, am^{1}, am^{6} and his-1 are restriction site differences and a superscripted + indicates the presence of that site. The observed number of each prototroph type is shown to the left of the bars. The Local Crossover group have an exchange between HP and HD while the Remote Crossover group have an exchange in either the sp to HP or HD to his-1

intervals. The Parental group have a parental association of sp and his-1.

Detailed Description of the Invention

Methods for Diversification of DNA

This invention provides a means to diversify heterologous DNA sequences *in vivo* to yield strains able to stably express the diversified genes directly. As used herein diversification refers to introducing sequence variation or new sequences into a DNA molecule or segment using the processes of recombination, crossing over, and/or conversion. In a preferred embodiment, the invention employs *cog*, a particularly active recombination hotspot in *N. crassa*, which generates discontinuous conversion tracts at high frequency. In consequence, nearby sites of sequence mismatch in the parent molecules are often not co-converted (see Figure 6) and the sequence of bases in the DNA of their progeny often has a new combination of the sequence differences that distinguished the parent molecules. The high frequency of discontinuous conversion tracts allow relatively infrequent events to be selected as a source of new variant sequences.

Diversified sequences are generated by constructing two interfertile (fertile) strains, parent (variant 1) and parent (variant 2), each having a different version of the gene or other foreign DNA sequence to be diversified juxtaposed to a recombination hotspot. The two strains are mated and the foreign sequences are diversified as a result of gene conversion, crossing over and/or errors in recombination that occur during meiosis as a result of the activity of the recombination hotspot. The foreign or heterologous DNA sequence is functionally coupled to the recombination hot spot. That is, the DNA sequence is in a position subject to recombination, crossing over, and/or conversion in the presence of the recombination hot spot. Where it is advantageous, the high frequency of discontinuous conversion tracts enhances of the yield of recombinants by selecting for progeny that have experienced a recombination event in a gene more distant from the recombinator than the foreign DNA to be diversified, ensuring that the whole of the foreign DNA sequence was covered by a conversion tract.

Where needed for the expression of heteromeric proteins, such as but not limited to immunoglobulins, these strains are also able to be combined in pairwise or other

combinations to form a panel of heterokaryons each of which stably express one of the possible combinations of the *in vivo* diversified foreign sequences.

The present invention provides a method for diversification of heterologous DNA molecules *in vivo* to generate new sequences that is advantageous and economical compared to previous methods. Diversification *in vivo* refers to diversification of foreign sequences after they have been introduced into living cells. This method significantly reduces the number of transfections that need to be performed and the number of genetically altered cells that must be checked for proper sequence insertion. The method of the invention generates a panel of different sequence variants, each being correctly integrated in the DNA of the host cell and thus able to be correctly expressed.

The reduction in the number of transfections required can be dramatic, particularly for monomeric or homomultimeric proteins. For example, if it is desired to generate 1024 new variant genes each coding for a version of a specific protein, *in vitro* diversification methods require 1024 transfections. Then each of the 1024 resultant transformed cells must be checked for correct insertion of the DNA (Figure 1A). However, if the 1024 new variants are generated using the present invention, only two transfections are required. This is a dramatic and advantageous reduction, a 512-fold reduction, in both the number of transfections that must be done and the number of transformants that must be checked for correct insertion of the heterologous DNA (Figure 1B).

The present invention can also yield a dramatic reduction in the number of transfections required for heteromultimeric proteins such as immunoglobulins. For example, new immunoglobulin variants can be generated by diversifying the immunoglobulin heavy chain gene, the immunoglobulin light chain gene, or both, and then combining them in pairwise combinations of a single heavy chain gene and a single light chain gene in a cell line able to express the genes correctly. 1024 new combinations of heavy and light chain genes can be obtained by combining 32 heavy chain variants with 32 light chain variants in all pairwise combinations. To generate the 1024 combinations using *in vitro* generated variant DNA sequences requires two transfections and checking the transformed cells for correct insertion of both a heavy chain variant and a light chain variant in each of 1024 cells (Figure 2A). This requires 2048 sets of complex

manipulations. This is a large number of complex manipulations.

Compared to *in vitro* methods, the number of transfections and manipulations can be reduced by using the heterokaryon technology described in US Patent Serial No. 5,643,745, issued to W.D. Stewart on July 1 1997. The Stewart method combines nuclei from a cell line expressing a heavy chain variant with nuclei from a cell line expressing a light chain variant to form a cell line containing nuclei of both types. In the Stewart method, it is necessary only to establish 32 cell lines carrying heavy chain gene variants and 32 cell lines carrying light chain gene variants to permit fusion of the cell lines into the 1024 different heterokaryotic cell lines having all possible pairwise combinations of one heavy chain gene with one light chain gene (Figure 2B). That approach requires only 64 transfections and subsequent checking for correct insertion of the variant DNA, a 32-fold reduction in the number of transfections and construct checking required, compared to previous *in vitro* methods.

The present invention provides a further sixteen fold reduction in the number of transfections that need to be done and constructs that need to be checked for correct insertion of the DNA compared to the Stewart method. The present invention requires only four transfections and only four constructs need to be checked for correct insertion to generate 32 new variants of light chain genes and 32 new variants of heavy chain genes required to test 1024 combinations (Figure 2C).

The larger the panel of variants to be tested, the greater the advantage of the present invention for *in vivo* diversification of DNA sequences. The limitation of new combinations that can be generated by the present invention is determined by the number of sequence differences between the DNA sequences recombined. It is known that 3.5% sequence difference over 3,000 bp is tolerated by the recombination mechanism. This is sufficient to generate about 10^{166} new variants from a single pair of sequences by recombination alone and an indeterminate large number more if errors occur in the recombination event. By using additional pairs of homologous DNA sequences (DNA sequences having substantial similarity but differing in one or more positions), an essentially unlimited variety of new sequence variants can efficiently be generated.

Conducting and Measuring Genetic Recombination Employing Fungi

The principal phase of the life cycle of fungi is haploid, making them advantageous for the study of recombination and for the generation of new genes or DNA sequences by the process of recombination or by errors in that process. Many fungi are known in the art to be suitable hosts for diversification of DNA, and filamentous fungi are preferred fungi. The invention includes fungal cells, either diploid or haploid and either as individual cells or as part of a larger assembly or organism, such as in macrofungi, and the like. Preferred strains of fungi used in *in vivo* diversification have several advantageous characteristics. Diversification is accomplished by mating a pair of haploid strains to create diploid parents, which then are crossed and cause *in vivo* diversification.

The haploid strains used to make the parents are of different and complimentary mating types. For example, a preferred pair of mating types is mating type *A. N. crassa* and mating type *a. N. crassa*. Fungi suitable for the invention also have other genetic characteristics required for heterokaryon compatibility. In *N. crassa* there are about eleven known loci that determine heterokaryon compatibility, *het-c, het-d, het-e, het-i, het-5, het-6, het-7, het-8, het-9, het-10 and mating type*. Preferred *N. crassa* carry the same allele of each of these loci except for mating types where one must have A mating type and the other a mating type. Preferably, fungi of the invention carry the same allele at each of the heterokaryon compatibility loci, except for mating type, to allow progeny of crosses to form heterokaryons in any combination of like mating type. Advantageously, a pair of fungi are used which include genetic features suppressing heterokaryon incompatibility between strains of different mating type. Such a genetic trait allows all combinations of progeny to form heterokaryons. Preferred *N. crassa* strains carry the mutation *tol*, which has been described by D. L. Newmeyer in Can. J. Genet. Cytol. 12:914-926 (1970).

Fungi of the invention can also include a forcing marker for the heterokaryon, such as one or more auxotrophic mutations. Forcing markers are preferred when the heterologous DNA codes for a component of a protein with more than one subunit. Preferably, if the multisubunit protein is made of different subunits, each parent will include the same forcing marker. If the multimeric protein has identical subunits,

preferably, each parent has a different forcing marker. Preferred forcing markers include mutations that inactivate one of the following genes: *trp-2, pan-2, thi,* or *arg.* Inactivation of these genes leads to a requirement for tryptophan, pantothenic acid, thiamine, or arginine, respectively.

For diversifying heterologous DNA, the heterologous sequence is inserted into the fungal genome between a recombinator (recombination hotspot) and a detectable gene, to create a recombinant fungal genome. Preferably, each of the parent strains carries an auxotrophic mutation. Preferably, when selection for recombination at the detectable gene is used to enhance yield of heterologous sequences diversified by recombination, the detectable genes in each parent strain will be alleles that do not complement to ensure that any rare aneuploid progeny (progeny having two copies of the chromosome carrying the detectable gene) cannot give rise to a heterokaryon carrying both alleles that will no longer be auxotrophic for the trait of the detectable gene and thereby falsely mimic the desired recombinants. *N. crassa* employed in the present invention preferably carry an *auxotrophic* mutant of the *his-3* gene. Preferably, each allele chosen has a mutation toward the 3' end of the *his-3* gene. A suitable non complementing allele pair is K26 and K480. Suitable complementing pairs of *his-3* alleles are K26 and K458 or K480 and K458. Suitable methods for selection of auxotrophic strains are described hereinbelow.

Advantageously, recombination at *his-3* can be used to select for progeny in which the yield of recombinants is enhanced by ensuring that the whole of the heterologous DNA sequence was covered by a conversion tract. Typically, such recombination can be carried out with the *a* and the *A* strains each carrying a different *his-3* allele and the SC medium will contain histidine in addition to any other required supplement. Preferably, the *his-3* alleles chosen are a non-complementing pair, for example, one will be K26 and the other K480 (Figure 10). This ensures that rare heterokaryons arising by the breakdown of anuploids containing two copies of chromosome 1 that form when there is failure of chromosome disjunction in meiosis, will not mimic his^+ recombinants. If the alleles were a complementing pair such as K26 and K458 or K480 and K458, the resulting heterokaryon would grow on medium lacking histidine, mimicing a his^+ recombinant. Suitable methods for selection of auxotrophic strains are described hereinbelow.

The strain is also chosen with characteristics that provide an active recombinator after insertion of the heterologous DNA into the strain. Preferably, each parent strain has a recombinator sequence and sequences surrounding the region in which heterologous DNA is inserted that have maximum homology to those regions on the plasmid. In preferred *N. crassa* parent strains, both cog^L and *lpl* sequences are from the Lindegren strain. Preferred fungi also include a regulator of the recombinator or recombination hot spots that allows or activates the recombination hot spot, which directs recombination to a particular locus on the gene. Preferred *N. crassa* strains carry *rec-2*, which directs cog^L to cause recombination in *his-3* and the inserted heterologous DNA.

Typically, each parent strain contains one or more genes conferring resistance to an agent used to select against the presence of the whole plasmid. Preferably, each strain includes the *mtr* gene which confers sensitivity to p-flurophenylalanine as a negative selective agent. The parent fungal strains can also include a genetic characteristic that limits their growth on plating media. Preferred *N. crassa* include the mutation *cot-1* C102t to limit growth on plating media. This strain can also include genetic characteristics such as mutations for additional sequences as required for enhance production and secretion of proteins produced by the heterologous gene or DNA sequence that is to be diversified. Some such genetic characteristics are described in U.S. Patent No. 5,643,745.

The life cycle of the filamentous fungus *Neurospora crassa* is outlined in Figure 4 (see also JRS Fincham, Genetics, Wright 1983). Two strains can form a dikaryon providing they have a different mating type. Mating type in fungal species is conferred by one or more types of mating type loci that carry short sections of idiomorphic DNA (NP Beatty *et al* Mycological Res. 98:1309-1316 1994). When two strains of different mating type are mixed together on an appropriate medium and incubated under appropriate conditions, dikaryons are formed and a developmental process is initiated that establishes a diploid cell that then undergoes meiosis, with the accompanying genetic recombination. The haploid products of meiosis are spores, each containing one of the four products of a meiotic division or in eight spored fungi such as *N. crassa*, four spore pairs each formed as a result of meiosis being followed by a mitotic division prior to spore formation.

Strains of *N. crassa* can be cultured on any medium suitable for this type of filamentous fungus containing a suitable carbon source, any required vitamins, and any nutritional requirements, for example, such as for an auxotrophic mutant. One such medium is that described by H. J. Vogel in American Naturalist **98**:435-446 (1964). The growth medium is solidified with agar when required. Typical *N. crassa* strains can be grown at temperatures between 20°C and 36°C. Typical temperature sensitive mutants are grown at either 25°C or 34°C depending on the type of growth preferred.

The *N. crassa* is cultured until asexual spores called conidia are produced on the aerial hyphae (Figure 4). The conidia are collected and used to propagate genetically identical cultures. Any suitable aconidiate mutants can be propagated by transfer of mycelial fragments to fresh medium.

Crosses of *N. crassa* are made by coinoculation of two strains of different mating type in a suitable medium for mating. Media suitable for mating include that of Westergaard and Mitchell (Am. J. Bot. **34**:573-557 1947) or SC medium (Davis et al. Methods in Enzymol. **17A**:79-143 1970), or Commeal medium. The crossing medium includes any specific nutrients required by the strains. The crossing medium is then incubated at a temperature suitable for the sexual cycle of *N. crassa*, typically 25 °C. Crosses of *N. crassa* can also be made by inducing a single strain to develop protoperithecia such as by incubation on low nitrogen medium for a suitable time and at a suitable temperature. Protoperithecia are fertilized by conidia of the second strain.

Fertilization leads to formation of a dikaryon, karyogamy, formation of diploid cells, meiosis and formation of ascospores. Typically, a convenient laboratory scale cross yields at least about 5×10^7 ascospores, or enough material to yield about 10^2 to about 10^5 or more, variants of heterologous DNA. Progeny from a cross are grown from ascospore by any appropriate method and individual progeny are recovered.

When one parent of a cross carries an auxotrophic mutation, ie a mutation requiring specific supplements to be added to the growth medium to compensate for their inability to synthesize the substance, half of the progeny of the cross will carry that mutation and will be unable to grow unless the growth medium is suitably supplemented. If both parents carry different auxotrophic mutations, only those progeny where the

mutant genes are recombined will be able to grow if no supplement is added to the medium. When the auxotrophic mutations are alleles (mutant forms of the same gene) and the differences from the wild-type are not at the same place in the sequence of bases that make-up the gene, prototrophic recombinant progeny (recombinants having the DNA sequence of the wild-type and thus able to grow without the supplement required by the mutants) occur at a low frequency, typically 0.001% to 1.0% of the total progeny and sometimes more, depending upon the gene, the distance apart of the sites of mutation and proximity to a recombination hotspot.

Neurospora Recombination Hotspot *cog*

The recombination hotspot *cog* in Neurospora is particularly suitable for *in vivo* diversification of foreign sequences. It is located approximately 4 kilobases distal of the *his-3* gene in which it stimulates recombination during meiosis (PJ Yeadon and DEA Catcheside Cur. Genet. 28: 155-163 1995). This strongly active hotspot, *cog*, has been located on the *N. crassa* genome, cloned (PJ Yeadon and DEA Catcheside Current Genetics, 28: 155-163 1995) and sequenced.

The relative placement of *cog* and *his-3* varies to some extent in natural strains (Yeadon and Catcheside *ibid*). The multiple sequence differences that distinguish two strains of *N. crassa* were used to map the location of *cog* with respect to the flanking genes *his-3* and *lpl* (Figure 5) and the position of exchange points of recombination events initiated by *cog*. The *cog* recombination hotspot is approximately 4 kb 3' of the *his-3* gene in the regulatory region of the *lpl* gene. In crosses between two different *his-3* alleles, the alleles are recombined in up to 1% of progeny. Such recombinants acquire tracts of sequence information from a homologous chromosome that can stretch at least 6 kb back to *cog*. These tracts are frequently interrupted (Figure 6). As a result, most progeny that experienced a recombination event initiated by *cog* have novel combinations of the polymorphic DNA sequences that distinguish the parents in the region between *his-3* and *cog*.

Recombination is easily measured in crosses between *his-3* mutants since those mutants require histidine for growth and recombination events that re-establish

prototrophy (*his*$^+$) permit the resulting strain to grow on media that do not contain histidine (T Angel *et al* Aust. J. Biol. Sci. **23:** 1229-1240 1970). Selection of prototrophs by their ability to grow without the addition of histidine to the medium allows the isolation of a subset of the progeny of the cross in which the DNA located between *his-3* and *cog* has experienced involvement in the intermolecular exchanges between homologous chromosomes that constitute a meiotic recombination event. Where the *his-3* mutants each had juxtaposed, between them and *cog,* different versions of a heterologous gene or other DNA sequence that differ at two or more sites, each *his*$^+$ recombinant will have a high probability of having a novel variant DNA sequence for the heterologous DNA due to recombination or to mutation, if errors occurred in repairing the DNA strands broken during recombination.

cog is an advantageous recombination hotspot for the purpose of *in vivo* diversification of heterologous DNA. This is because it catalyses a high frequency of recombination, generates long conversion tracts that are frequently interrupted, recombination in its vicinity is known to be tolerant of high levels of DNA sequence polymorphism, and the nearby *his-3* gene has complementing alleles allowing the targeting of constructs into the *his-3* gene without loss of a *his-3* mutant marker required for enrichment of progeny in which the juxtaposed heterologous DNA has been diversified by recombination. Heterologous DNA to be diversified is functionally coupled, or juxtaposed, to the recombination hot spot. That is, the heterologous DNA is in a position in the genome, plasmid, or other DNA sequence subject to recombination in the presence of the recombinator or recombination hot spot. The degree of divergence in normal chromosomal sequence that has been used to study recombination in yeast is more than an order of magnitude less than the 3.5% sequence divergence in the 3' flank of *his-3* used as markers in the study of recombination catalyzed by *cog* (Figure 5, Figure 6). Elimination of heterology in the 3' flank of *cog* led to less than two fold increase in gene conversion in *his-3*, demonstrating that the Neurospora *cog* recombinator is tolerant of the high levels of sequence heterology required for efficient sequence diversification *in vivo* by recombination of homologous DNA having many sequence differences. Two alleles of *cog* are known. *cog*^{L}, present in the Lindegren laboratory strain, is a preferred

recombinator. cog^E, present in the Emerson and St Lawrence 74A laboratory strains, yields recombination frequencies an order of magnitude lower than does cog^L.

The Neurospora cog hotspot is preferred due to advantageous characteristics such as a high frequency of interrupted conversion tracts, suitable marker genes for targeting plasmid constructs and enhancing the yield of recombinant forms of heterologous DNA, and promoters suitable for expressing the diversified sequences and secreting any protein product. The cog locus is a particularly preferred recombination hot spot. However other hot spots with advantageous characteristics of the cog locus are also preferred.

Other recombination hotspots in Neurospora and in other species are also suitable for *in vivo* diversification. Indeed, recombination hotspots can be expected in all genomes of those eukaryotes that reproduce sexually and also in the genomes of bacteria and other prokaryotes and their plasmid and viral parasites that are able to exchange genes. Such recombination hotspots include ones 5' of *his-3* and 5' of *am* in N. crassa, 5' of each of *his4*, *arg4*, *his2*, and *cys3* in S. cerevisiae, both within and 5' of *ade6* and in *S. pombe*, and the like. Additional suitable hot spots include hotspots that affect recombination at *his-1* and *nit-2*, near *pyr-3*, near *sn*, and near *his-2* in *Neurospora crassa* and hotspots known to exist in the fungi *Aspergillus nidulans*, *Schizophillum commune*, and at other locations, including HOT1 in the genome of *Saccharomyces cereviseae*. Among these other suitable hot spots, ones 3' of *his-3* and 3' of *am* in *N. crassa*, 3' of *his4* and 3' of *arg4* in *S. cerevisiae*, within *ade6* in *S. pombe*, and the like, are believed to be more suitable.

Hot spots are also found in plants, such as maize. Mammalian hot spots include those in mouse (Mus musculus), where recombination hotspots are known close to the major histocompatibility locus, in human (Homo sapiens) and Chimpanzee, where hotspots are known near the gamma globulin loci, and, in humans, also near the retinoic acid alpha receptor gene and in the region of the repeat sequences associated with Charcot-Marie-Tooth neuropathy. There are also suitable recombination hot spots in diverse species.

Replacement of Sections of Chromosomes with Altered or Heterologous DNA

DNA sequences in fungal chromosomes can be altered by incorporating sequences

that have been engineered *in vitro* in place of those in the chromosome. The engineered sequences can contain heterologous DNA in addition to altered forms of sections of the chromosomal DNA and can be additional to or a replacement for sections of chromosomal DNA sequences. The means of making these manipulations utilize the methods known in the art for DNA manipulation including: oligonucleotide synthesis, PCR amplification of DNA, restriction enzyme digestion, cloning in appropriate phage or plasmid vectors using suitable hosts species such as *Eschericia coli* as outlined in standard methods manuals (for example J Sambrook *et al* "Molecular Cloning" Cold Spring Harbor 1989), and the like.

The *in vivo* constructs can be introduced into fungal cells by a variety of procedures, referred to herein as transfection, that are known in the field of DNA manipulation (J Sambrook *et al* 1989, FJ Bowring and DEA Catcheside Current Genetics 23: 496-500 1993) such that the *in vitro* modified DNA takes the place of the normal chromosomal sequence. Precise replacement requires that the DNA is incorporated in exactly the correct position by a breakage of the chromosomal DNA and insertion of the *in vitro* construct. In many species, exactly homologous insertion events are rare and selection of correct insertions is required to avoid DNA insertion at ectopic sites with limited or no DNA sequence homology. Correctly targeted insertions can be achieved by repair of a chromosomal mutation. Transplacement can be achieved as follows.

A DNA construct is prepared containing a shortened version of the *his-3* gene from *Neurospora crassa* truncated at the 5' end and extending some distance 3' of the gene. The heterologous or *in vitro* modified DNA sequence to be targeted into the recipient cells inserted into this DNA at some point 3' of the *his-3* gene. The exact junction for the point of insertion of the heterologous DNA can vary from construct to construct and be effected at a multiple cloning site built into the sequences 3' of the terminus of the *his-3* gene, or whatever other gene is being targeted. The requirement is that the junction is not in DNA sequences coding for an essential cellular function.

Suitable plasmids include pBM60 (Margolin *et al* Fungal Genet. Newsl. 44:34-35 1997), pRAUW122 (R Aracano and RL Metzenberg Fungal Genet. Newsl. 43:9-13 1995), and pFJB1 (FJ Bowring and DEA Catcheside Current Genet. 23: 496-500 1993), which are capable of replicating in *Escherichia coli* cells. Such a plasmid typically contains a

insert blunt 2 here

truncated *his-3* gene of *N. crassa*, sequences 3' of *his-3* into which have been inserted a selectable marker, such as the *hph* gene coupled to a promoter that permits expression in *N. crassa.* The marker gene thus confers resistance to the antibiotic *hygromycin.* The plasmid also contains a multiple cloning site for insertion of heterologous genes. The plasmids typically also contain a selectable marker such as *bla,* which confers ampicillin resistance to *E. coli* cells containing the plasmid. The plasmid also typically contains an *E. coli* DNA replication origin, permitting selection for cells containing the plasmid and cloning in *E. coli* of plasmid DNA into which heterologous sequences have been inserted into the multiple cloning site. To be suitable for the present invention, the plasmid lacks substantial sequence differences in the *his-3* gene from that in the Neuropsora strain into which it is to be targeted and preferably contains the high frequency recombinator cog^{L}.

A suitable plasmid can be constructed using methods, in the art for DNA manipulation. A preferred plasmid includes a series of DNA sequences in the following order. First, the preferred plasmid has a sequence including the majority of the *his-3* gene, but lacking a portion of the 5' end of the gene. Preferably, the *his-3* gene lacks a short portion of the sequence of the 5' end including the start codon. Preferably, the excluded sequence includes about 300 nucleotides from a position not less than about 687, preferably to about position 1000 in the sequence shown in Figure 7. This first sequence preferably terminates beyond the stop codon of the *his-3* gene. Preferably, the first sequence is a sequence from a $his\text{-}3^{+}$ strain of *N. crassa*, preferably the St. Lawrence 74A wild type, the Lindegren wild type, mutant K26 derived from Lindegren, or K458 mutant derived from Emerson A wild type.

Second, is an optional sequence including a promoter that is functional in *N. crassa*. Preferably, the second sequence is included when no such promoter is found in other DNA in the plasmid. Preferably the optional promoter sequence is functionally coupled to the expression and control of expression of a heterologous sequence.

Third, comes another optional sequence. The third sequence can be present when the heterologous or foreign sequence codes for a messenger RNA. When such a messenger RNA is produced, the fourth sequence can tag the messenger RNA for export of the protein product. Preferably, the third sequence is included when such export is

desired and when no suitable tagging sequences are present elsewhere in the plasmid, such as in the heterologous or foreign DNA insert.

Fourth, is a sequence including a cloning site having cleavage sites for one or more restriction enzymes. This cloning site can be any suitable site for insertion of a heterologous gene or DNA sequence, preferably a gene or sequence not present elsewhere in the plasmid.

Fifth, is a recombination hotspot, or recombinator, sequence. Preferably this is a *cog* sequence, preferably a cog^L allele. A preferred *cog* allele is from the 3' of the Lindegren wild type, preferably from about nucleotide 5412 to about nucleotide 6831.

Sixth, comes a sequence that provides homology downstream of *cog*. Preferably the sixth sequence includes sequences from the *lpl* gene, preferably from position about 6831 and in the 3' direction for about several hundred base pairs in the sequence shown in Figure 7.

Seventh, is a sequence providing a marker gene providing for either positive or negative selection for the presence of the plasmid in *N. crassa*. A preferred marker gene for positive selection is hph^R or another gene conferring hygromycin resistance. A preferred negative selection gene is mtr^+, conferring sensitivity to p-flurophenylalanine. Additional suitable genes providing for negative selection include those conferring sensitivity to a toxic substance.

Eighth is a sequence that provides a maker gene for the presence of the plasmid in *E. coli*. A preferred eighth sequence includes the *bla* marker gene or another gene encoding ampicillin resistance. Additional suitable genes include those conferring resistance to another antibiotic.

Ninth is a sequence that provides for amplification of the plasmid in *E. coli*. Preferably the ninth sequence is a replication origin functional in *E. Coli*.

Variants of this preferred plasmid can be constructed as well. Such variants include a plasmid lacking the sixth sequence, but relying on another sequence for homology downstream of *cog*. Such homology can result from the cog^L sequences, which are included in the fifth sequence.

The construct is introduced by methods such as transfection, transformation or

electroporation into a *N. crassa* cell or protoplast or spheroplast that contains a mutation in the gene to be targeted. Typically the targeted gene is the *his-3* gene and the mutation is located some distance 3' of the point of truncation of the *his-3* sequences within the DNA construct. The mutation renders the gene product non-functional, requiring that growth conditions are modified to permit cell growth. Typically, by addition of an amino acid such as histidine or other supplement as is required to the growth medium.

In this embodiment, the recipient cells are placed on medium lacking histidine to select for those that have had the histidine gene restored to functionality by insertion of the DNA construct or by a mutation event restoring a functional *his-3* gene. Cells in which mutation events rather than transplacement events occur can be rejected by incorporating a selectable marker such as the *hph* gene that confers resistance to the antibiotic *hygromycin* as part of the heterologous DNA, using recipient cells that are hygromycin sensitive, and culturing them on medium containing hygromycin as well as lacking histidine.

The specific selectable marker built into the *in vitro* modified DNA can vary and is not confined to hygromycin resistance, or, indeed, to antibiotics. Alternative markers include a gene for synthesis of an essential metabolite where the recipient cell has that essential gene deleted. Cells that are able to grow on the selective medium will include those where exchange between the *in vitro* modified DNA and the cellular DNA has occurred in the interval between the 5' terminus of the truncation of the *his-3* DNA sequences and the site of the *his-3* mutation in the chromosome. This will restore a complete active copy of the *his-3* gene in the chromosome with all or part of the *in vitro* construct, including any heterologous DNA built into it, inserted into the chromosome specifically at the *his-3* gene.

Where only a single reciprocal exchange event occurs, the heterologous DNA will be flanked each side by DNA sequences that are homologous to one another. Each of these two DNA sequences is a near duplicate of the other comprising the *his-3* gene and flanking sequences from the *in vitro* construct and the corresponding chromosomal copy. These sequences will be reciprocally recombined in a manner determined by the precise location of the single reciprocal exchange event.

Where exchange events occur both in the region of homology within *his-3* and in

the region of homology in the 3' flank of *his-3*, there will be no duplication and a proportion of the *in vitro* construct DNA including the whole of the heterologous sequence will be inserted exactly in place of the normal chromosomal DNA sequence in this region. This is termed transplacement. Transplacement can be effected in one step or in two steps in which the first exchange leads to insertion and the generation of a duplication and the second exchange leads to resolution of the duplication leaving the *in vitro* construct in place of the excised normal chromosomal sequence. Confirmation of correct transplacement can be accomplished by standard methods of DNA manipulation using techniques such as southern transfer, or PCR amplification and restriction enzyme digestion, to test the architecture of the relevant chromosomal region of the selected cells.

Duplicated DNA in some organisms, including *N. crassa*, suffers extensive mutations as a result of conversion of cytosine bases to thymine in a process known as repeat induced point mutation (RIP) that occurs during the expansion of the dikaryotic tissue which precedes karyogamy and meiosis (EU Selker Ann. Rev. Genet. 24: 579-613 1990). The mutations are so extensive that genes within the duplication are usually inactivated, making it unsuitable as a gene diversification mechanism. It is therefore advantageous to avoid duplications of a size susceptible to RIP in any gene manipulation process that requires passage of DNA through meiosis.

Constructing Strains of Fungi

The invention also relates to a method of preparing a strain of a fungus. Advantageously such a strain of fungus can be used for producing diversified DNA sequences and the gene products of the diversified sequences. Such a strain can provide markers for the isolation of recombinant forms of heterologous DNA, and production and isolation of diversified DNA. This method requires making a diploid fungal cell containing non-complementary alleles of a gene providing an auxotrophic mutation, which gene is functionally coupled to, adjacent to, or juxtaposed to the heterologous DNA to be diversified. This diploid allows enrichment for cells containing diversified DNA. A typical method used for making such a strain of fungus includes several steps.

A first fungal cell having a first allele of the gene providing an auxotrophic

mutation is transfected with a first heterologous DNA and a second allele of the gene providing an auxotrophic mutation. In this method transfection can be accomplished employing a suitable cloning vehicle such as a plasmid, a viral vector, another suitable vector, and the like that includes the heterologous DNA and the desired allele of the gene. Advantageously, each of the first and second alleles encode a defective gene but are complementary alleles. The presence of the first and second allele of a gene following transfection of the first fungal cell establishes a first heterokaryon, which is grown to provide a first homokaryon containing the second allele of the gene and the first heterologous DNA.

A second fungal cell having a third allele of the gene providing an auxotrophic mutation is transfected with a cloning vehicle including a second heterologous DNA and a fourth allele of the gene providing an auxotrophic mutation. Advantageously, each of the third and fourth alleles encode a defective gene but are complementary alleles. The presence of the third and fourth allele of the gene providing an auxotrophic mutation following transfection of the second fungal cell establishes a second heterokaryon, which is grown to provide a second homokaryon containing the fourth allele of a gene and the second heterologous DNA. In a preferred embodiment, the first allele and the third allele are the same allele and the second and fourth alleles of the gene are non-complementing alleles. The desired fungal strain is established by crossing the first and second homokaryons.

Advantageously, in this method, the fungus is *Neurospora crassa*. Preferably, in *Neurospora crassa* the auxotrophic mutant is a *his-3* auxotrophic mutant. Preferred auxotrophic mutants include *his-3* K26, *his-3* K458, and *his-3* K480. Preferably, these are used in combinations such that the first and second fungal cells each carry *his-3* K458, and the first fungal cell is transfected with his-3 K26 and the second transfected fungal cell is transfected with *his-3* K480. Advantageously, the non-complementing pair of alleles is K26 and K480, which results in production of a dikaryon carrying both alleles that is unable to grow on media lacking histidine.

More specifically, targeting plasmid DNA into a specific chromosomal location is achieved by transfection of an auxotrophic mutant with a plasmid carrying non mutant

sequence leading to the restoration of a normal gene in the recipient chromosome (Figure 9) or with a plasmid carrying a complementing mutant gene (Figure 10). Transformants with the plasmid DNA correctly targeted are selected by their ability to grow on media not supplemented with the requirements of the auxotrophic mutation. Where the fraction of progeny from a cross that experienced conversion of heterologous DNA is too low to provide a good yield of diversified sequences, it is necessary to be able to enrich for those that did experience conversion to provide the panel of diversified sequences. This can be achieved if a *his-3* mutation remains in the chromosome of the recipient cell after insertion of the heterologous DNA.

In *Neurospora crassa*, several auxotrophic markers can be used for isolation of desired heterologous DNA. One preferred marker is the *his-3* gene. Complementing allelic mutations of the *his-3* gene (DG Catcheside and T Angel Aust. J. Biol. Sci. 27:219-29 1974) provide a means of achieving transplacement at *his-3* leaving a *his-3* mutation conferring a requirement in the chromosome carrying the heterologous DNA. This is made possible because *his-3* codes for a peptide that forms a heteromultimeric protein, which allows selection of strains producing different combinations of subunits. Certain combinations of subunits will be inactive, and others active. For example, an enzyme formed by subunits coded by *his-3* K458 is inactive and a heteromultimer containing subunits coded by *his-3* K26 and other subunits coded by *his-3* K458 is enzymatically active. This allows selection of heterokaryons that contain both types of nuclei. The desired heterokaryon can be specifically selected by its ability to grow on media that contain no added histidine.

A homokaryon containing only the desired nuclear type, for example, carrying the desired *his-3* mutation and preferably the juxtaposed heterologous DNA, is isolated by picking colonies to slopes of minimal medium, growing to conidia and isolating homokaryons by establishing new cultures on medium containing histidine from single conidia. The conidia frequently give rise to homokaryons. The homokaryon can then be used as one parent of a cross used to diversify the heterologous DNA (Figure 10). A second homokaryon having a non-complementary allele of the gene providing auxotrophy is prepared analogously to the first homokaryon. This provides the second parent for

crossing.

The present invention can be better understood with reference to the following examples. These examples are intended to be representative of specific embodiments of the invention, and are not intended as limiting the scope of the invention.

Examples

Example 1 -- Long, Interrupted Conversion Tracts by *cog* in *Neurospora crassa*

We have used 14 PCR products ranging from 330 to 540 bp in length each having easily detectable restriction site polymorphism (RSP) or sequence length polymorphism (SLP) to investigate the molecular outcome of conversion within *his-3* and 3.8 kb distal of this gene. A *Hpa*I RSP allowed extension of the analysis to 6 kb proximal of *his-3*. The RSPs and SLP s were used to determine the parental origin of each segment in 38 progeny prototrophic for histidine from crosses heteroallelic for *cog* and *his-3*. Progeny from diploids that were homozygous *rec-2* and heterozygous *rec-2/rec-2*$^{+}$ were examined to investigate differences in outcome due to the inactivation of the *cog* function by *rec-2*$^{+}$ and from diploids in which the *his-3* mutation closer to *cog* was *cis* (K1201/K26 and K504/K26) or *trans* (K26/K874) to *cog*^{L} to detect differences resulting from the bias for conversion to be initiated on the *cog*^{L} chromosome. The large number of heterologies scored (16 in 6.9 kb, including the sequence variations responsible for the *his-3* mutations) permits measurement of the length of conversion tracts and detection of discontinuity and location of crossovers between flanking markers *arginine-1* (proximal of *his*-3) and *ad-3* (distal), if they occur within the region surveyed.

Materials and Methods

Origin of Neurospora parental strains: K26 (Table 1) was isolated in a strain of Lindegren 25*a* origin (Catcheside and Angel 1974 Aust. J. Biol. Sci. 27: 219-229.); K1201, K504, K374, *arg-1* K166 and *ad-3* K118 in Emerson *a* (Catcheside and Angel 1974 *ibid*). T10988 and T10990 were generated by Steve Fitter and T10997 by Fred Bowring. F strains are from the collection of D.G. Catcheside.

Culture methods: Methods were those described by Bowring and Catcheside (1996 Genetics 143: 129-136.), except that crosses were supplemented with 200 µg/ml alanine, 500 µg/ml arginine and 200 µg/ml adenine. Vegetative cultures were supplemented with 500 µg/ml alanine, 500 µg/ml arginine, 200 µg/ml histidine, and 400 µg/ml adenosine as required.

Isolation of recombinant progeny: Ascospores were treated as described in Catcheside (1981 Genetics **98:** 55-76) except that plates contained 0.05% of glucose and fructose in place of sucrose were supplemented with arginine, adenosine and alanine and incubated for 3 days at 34°. *His*$^+$ colonies were picked to slopes and grown at 25°. Cultures were streaked for single colonies and reisolated before further analysis. Flanking markers were determined by the ability of a prototroph to grow either without adenosine or without arginine.

T11245-T11320 (Table 1) are histidine prototrophs isolated for the purpose of conversion tract mapping. T11245-T11252 were derived from a cross between T10988 and T4393, T11253-T11260 from a cross between T10990 and F7446, and T11261-T11268 from a cross between T10990 and F7448. T11269-T11274, T11302-T11305 and T11320 were derived from three separate crosses between T10987 and T4398. T11275-T11277 and T11306-T11308 were from two independent crosses between T10990 and T10997.

Preparation of PCR templates: Quick template DNA was made from each progeny strain as described in Yeadon and Catcheside (1996 Fungal Genetics Newsletter **43:** 71.). For the parental strains, genomic DNA was prepared as described by Yeadon and Catcheside (1995 Curr. Genet. 28: 155-163.).

PCR amplification: PCR was performed for 40 cycles (Saike, et al. 1988 Science 239: 487-491) using a PTC-100 Thermal Sequencer (MJ Research Inc., supplied by Bresatec) and *Taq* DNA polymerase (BTQ-1; Bresatec) in a total reaction volume of 50µl. Fifty nanograms of genomic DNA or 2 µl of quick template DNA was used as template in each reaction. Annealing was at 50° and $MgCl_2$ 2.5 mM.

PCR primers: *his-3* primers were designed using the program PCRPRIM on

ANGIS from the sequence of *histidine-3* published by Legerton and Yanofsky (1985 Gene 39: 129-140) and corrected as necessary where sequence information conflicted with the published sequence. Primers distal of *his-3* (Figure 5) were designed using sequence previously obtained for the intergenic regions (Yeadon and Catcheside 1995 Curr. Genet. 28: 155-163). The P1 pair of primers (Figure 5) was designed after a sequence walk proximal of *his-3*.

Restriction digests and electrophoresis: PCR products were digested with 3 units of the appropriate restriction enzyme (New England Biolabs) for 90 min. as described by the manufacturer and the products resolved by electrophoresis on 3% NuSieve 3:1 agarose (FMC Bioproducts) in TAE, 3 V/cm for 3 hr.

Detection of RSP and SLP in the Lindegren and Emerson parents: Genomic DNA from T10987 (Lindegren descent) and F7448 (Emerson descent) was used as template to amplify DNA segments (Figure 5) predicted from the sequence to have RSPs or SLPs that differentiate each parent. PCR products were digested where necessary with appropriate restriction enzymes and fragments resolved by electrophoresis (data not shown). Each pair of PCR products yielded the expected distinguishable patterns, length polymorphisms in GAP(X) and C8, and a restriction site present in only one parent in the remainder. The polymorphism in C8 reflects the presence of the inverted repeat transposable element *Guest* in Emerson (Yeadon and Catcheside 1995Mol. Gen. Genet. 247: 105-109) yielding a product 102 bp longer than that from Lindegren. The *Hpa*I RSP 6 kb proximal of *his-3*, discovered during mapping of genomic DNA (Yeadon and Catcheside 1995 Curr. Genet. 28: 155-163, Figure 5), assisted location of crossovers proximal of *his-3* and was detected by Southern analysis as described in Yeadon and Catcheside (1995 Curr. Genet. 28: 155-163.). The probe was λJY25 (Yeadon and Catcheside (1995 Curr. Genet. 28: 155-163.).

Determination of the parental origin of sequence segments from recombinant progeny: Fourteen of the 15 segments (Figure 5) were amplified by PCR from the 38 histidine prototropic progeny, and the parental origin of each identified from their restriction pattern or, in the case of C8, from the length polymorphism due to *Guest*.

The segment labeled GAP(X) (Figure 5) was not used since reliable sequence was difficult to obtain, impeding identification of RSPs and, since a single site could not be surveyed, the parental origin of segments different in length from that in either parent would be uncertain.

Results

Bias in the chromosome receiving information: Among the 23 progeny of crosses homozygous *rec-2*, four strains (T11247, T11265, T11266 and T11267) had exchanges with no detectable conversion (henceforth termed "simple crossovers") between the *his-3* mutations. In two others, T11262 and T11264, there were alternate positions in which the crossover could have occurred, indicated by the symbol "X?" in Figure 6. In 16 of these 23 progeny the Lindegren chromosome must be the recipient of information, and in T11261 alone was Emerson most probably the recipient of information. It is clear that the Lindegren chromosome, which carries the high frequency cog^L allele, was converted more often than the Emerson chromosome. In contrast, in the 11 progeny from crosses in which $rec\text{-}2^+$ is present, the two chromosomes were equally likely to be converted: one has a simple crossover (T11269), there were four in which Emerson was converted (T11270, T11274, T11307 and T11320), three where Lindegren was converted (T11275, T11276, and T11306) and three (T11272, T11302 and T11304) where the recipient of information cannot be determined.

Positions of crossovers: Nine of the 23 progeny from *rec-2/rec-2* crosses (39%) had at least one crossover between *arg-1* and *ad-3* (Figure 6). Three of these were between *arg-1* and the *Hpa*I RSP (H) 6 kb proximal of *his-3* (in T11264, T11266 and T11268), one was between H and *his-3* (in T11252), seven were within *his-3* (in T11245, T11247, T11262, T11264, T11265, T11266 and T11268) and one (T11261) was between *cog* and *ad-3*. Seven of the 11 progeny from crosses in which $rec\text{-}2^+$ was present (64%) had crossovers between the flanking markers (Figure 6). Two of these were between *arg-1* and H (T11276 and T11307), four were within the *his-3* gene (T11269, T11272, T11302 and T11304) and one (T11275) between the C3 and C9

heterologies distal of *cog*.

Length of conversion tracts: Conversion tracts in prototrophs from crosses homozygous *rec-2* vary in length (Figure 6). All prototrophs, except those with simple crossovers and the two potential revertants, showed conversion of more than one marker. The longest continuous tracts were those in T11254 and T11259 that cover the region between the R1 and C9 heterologies, 5.6 kb apart. RSP typing distal of C9 showed that the tract in T11254 terminates <185 bp distal of the heterology in C9, and that in T11259 ends >300 bp and <1178 bp distal of the C9 heterology (data not shown). Thus the longest tract extended >5.9 kb.

The presence of *rec-2*$^+$ results in shorter conversion tracts that did not extend distal of *his-3*. The longest continuous conversion tracts in these progeny are in T11270 and T11274. The tracts cover R1, K504 and P1 were thus at least 940 bp long.

Discontinuity in conversion tracts: Among progeny from crosses homozygous *rec-2*, there were 17 that showed evidence of conversion (Figure 6). Of these, conversion tracts were discontinuous in eight (47%). Ten of the 11 progeny from crosses including *rec-2*$^+$ (Figure 6) showed evidence of conversion and in three of these the tracts were discontinuous (30%). The difference between discontinuity of tracts in crosses in which *rec-2*$^+$ was present or absent was not significant ($X^2 = 0.25$ with Yates' correction; P>0.5). In total, 11 of 27 (41%) conversion tracts were discontinuous.

Association of crossovers with conversion: At least one crossover between the flanking markers *arg-1* and *ad-3* occurred in 47% of the prototrophs (16 of 34; Figure 6). Of the prototrophs from crosses homozygous *rec-2*, three of the crossovers (in T11264, T11266 and T11268) were between H and *arg-1* and were likely to be too distant to be associated with conversion. Those crossovers in T11252 and T11261 are also sufficiently distant that their association with conversion is doubtful. This leaves seven crossovers that may be associated with conversion initiated by *cog*^{L}, although four of these are simple crossovers, with no evidence of conversion. Thus, ignoring simple crossovers, 16% of prototrophs (three of 19) from crosses homozygous *rec-2* had crossovers apparently associated with conversion. If simple crossovers are included as potentially associated with conversion, then 30% of these prototrophs (seven of 23) had

an associated crossover.

Five of the seven crossovers in progeny from crosses heterozygous *rec-2/rec-2*$^+$ were within the region surveyed; one was a simple crossover (T11269), one (T11275) was sufficiently distant that an association with conversion is doubtful, but three (in T11272, T11302 and T11304) were at ends of conversion tracts and thus may be associated with conversion. The two remaining crossovers (in T11276 and T11307) were >6 kb from the proximal end of *his-3* and were unlikely to be associated with conversion at this locus. Of the 10 prototrophs from crosses heterozygous *rec-2/rec-2*$^+$ (*cog* inactive) that had evidence of conversion, three (30%) have crossovers that may be associated with conversion. If simple crossovers are included, 36% of these prototrophs (four of 11) had crossovers that may be associated with conversion.

In total, 11 of 34 prototrophs (32%) had crossovers that may be associated with conversion. The presence or absence of *rec-2*$^+$ had no significant effect on the association between conversion and crossing over in prototrophic progeny (X^2 = 0.31 with Yates' correction; P> 0.5).

Discussion

Regulation of recombination in the *his-3* region of Neurospora by the unlinked gene *rec-2*, the dominant allele of which (*rec-2*$^+$) prevents the initiation of recombination at *cog* (Catcheside and Angel 1974 Aust. J. Biol. Sci. **27:** 219-229.), allows separation of events at *his-3* into those that are *cog*-related and those that are not. Apart from the additional aspect of local regulation of recombination in Neurospora, there are many similarities between *cog* and yeast hotspots that suggest a common mode of activity.

Convertants at *his-3* manifest a bias in the direction of information transfer. Of the 23 prototrophs from crosses lacking *rec-2*$^+$, in only one was Emerson, almost certainly the recipient of information (Figure 6). Likewise in both budding and fission yeasts, where a hotspot has alleles that differ in activity such as the *ade6 M26* hotspot of *S. pombe* (Gutz 1971 Genetics **69:** 317-337) or the promoter deletion that removes hotspot

activity at *ARG4* of *S. cerevisias* (Nicholas, et al. 1989 Nature 338: 35-39), the chromosome on which recombination is initiated is the recipient of information.

Conversion tracts at different hotspots vary in length, dependent both on the locus and the mode of selection of progeny. The length of conversion tracts at *his-3* in Neurospora is within the range of those measured in yeast at several loci. Conversion tracts at *his-3* can be over 5.9 kb long (Figure 6), but the distance between the recombinator and the gene may select for a subset enriched for the longest tracts in prototrophic progeny. The degree of discontinuity of conversion tracts appears to vary between loci as well as between species. Forty-one percent of conversion tracts in this study are discontinuous (Figure 6). The discontinuities of conversion tracts occur within the intergenic and divergent sequences between *his-3* and *cog,* that have no known cellular function, further diversifying these sequences. This shows that any pair of homologous but divergent foreign DNA sequences inserted into this region will also be diversified.

Example 2 -- Analysis of Conversion Tracts Associated with Recombination Events at the *am* Locus of *Neurospora crassa*

Introduction

The incidence of crossing over between flanking markers is enhanced when gene conversion is observed at an intervening locus and this is taken as evidence that gene conversion and crossing over are intimately associated. This view was strengthened when Hurst *.et al*, (Hurst D, Fogel S, Mortimer R (1972) Conversion associated recombination in yeast Proc Natl Acad Sci USA 69: 101-105) reported that in *Saccharomyces cerevisiae*, half of conversion events enjoyed an associated crossover. However, these data were not corrected for incidental exchanges in the regions flanking the converted loci (Stadler D R (1973) The Mechanism of intragenic recombination Ann Rev Genet 7: 113-127) and when that correction was made, approximately 35% (r = 0.35) of conversion events were found to have an associated crossover (Fogel S, Mortimer R, Lusnak K, Tavares F (1979) Meiotic gene conversion: a signal of the basic recombination event in yeast Cold Spring Harbor Symp Quant Biol 43: 1325-1341).

The level of association between conversion and crossing over reported for Neurospora is similar to that in *S. cerevisiae* (Neurospora $r = 0.33$) and the Neurospora *am* locus is not extraordinary in this respect (*am*, $r = 0.26$; Perkins D D, Lande R, Stahl F W (1993) Estimates of the proportion of recombination intermediates that are resolved with crossing over in *Neurospora crassa* Genetics 133: 690-691).

Restriction site polymorphisms (RSPs) either side of and tightly linked to the Neurospora *am* locus, were used to examine the association between gene conversion and crossing over (Bowring and Catcheside, 1996). Analysis of prototrophs from a repulsion phase cross heteroallelic for the mutations am^1 and am^6 (cross B163) revealed that the majority of crossovers between conventional flanking markers were outside of the region bounded by RSPs and thus, that these events were remote from *am* and the event that generated a prototroph. It was concluded that a maximum of 7% ($r \leq 0.07$) of *am* conversions enjoyed an associated crossover.

am and flanking regions in the B163 parents were sequenced which revealed nine sequence polymorphisms. Five of these, together with the two mutant alleles have been used to investigate the nature of recombination events in B163 prototrophs in more detail.

Materials and methods.

Crossing methods, prototroph isolation, DNA preparation and classification of flanking markers were described in Bowring and Catcheside (1996)(Genetics 143: 129-136.) B163 is a cross between F11089 and F6325. F11089 (*A, rec-3; cot-1* C102(t); *sp* B132, am^1) and F6325 (*a, rec-3; cot-1* C102(t); am^6, *his-1* K627) are D R Smyth and D G Catcheside stocks respectively. Among 205 am prototrophs from B163 selected at random, 145 were sp^+ *his-1*, 14 were *sp* $his\text{-}1^+$, 16 were sp^+ $his\text{-}1^+$ and 30 were *sp his-1*. 84 of these (27 sp^+ *his-1* randomly selected, all of the sp^+ $his\text{-}1^+$ progeny, all but 1 of the *sp* $his\text{-}1^+$ and 2 of the *sp his-1* progeny lost during processing) were assayed for molecular markers (Bowring and Catcheside, 1996, *ibid*).

Primers (Figure 11) were designed using the program "PCRPRIM" on ANGIS

and the *am* sequence data of Kinnaird and Fincham (1983). Primer 0L, the upstream end of which is located 188bp beyond the upstream extent of Kinnaird and Fincham's sequence, was designed using sequence data obtained from J.A. Kinsey. PCR was carried out in a Corbett FTS-1 cycler. $MgCl_2$ was included at a final concentration of 2.5mM. TAQ polymerase was from Bresatec (Adelaide, South Australia). A five minute denaturation step was followed by 30 cycles with 1 min. at 94C, 1 min. at 50C and 1 min. at 72C.

PCR products were purified (Wizard, Promega) and sequenced on an Applied BioSystems automated sequencer using both forward and reverse primers. Digestion of PCR amplified fragments with the appropriate restriction enzyme was used to determine the alleles carried by prototrophs isolated from cross B163: the *F* polymorphism by digesting the PCR product amplified with the 0L-1R primers with *Fok*I; *Bd* and *Bp* by digesting the fragment amplified with 1L-2R with *MboI*; *S*, by digesting the sequence amplified with 1L-2R with *Mse*I and *A*, by digesting the fragment amplified with 2L-3R with *Bsm*A1. Primer 0L was used with 1R in the *F* determination since the polymorphic *Fok*I cleavage site was very close to primer 1L. In all determinations except *S*, the amplified fragment had at least one non-polymorphic site for the appropriate restriction enzyme allowing for detection of false negatives. While this was not possible for *S*, because the same PCR product was digested with *Mbo*II for the determination of *Bp* and *Bd* and *Mse*I for the determination of *S*, and as none of the products were refractory to *Mbo*II digestion (data not shown), it seems unlikely that any progeny were misclassified for *S*. Digestions were carried out according to the suppliers suggested protocol. *Mbo*II was from Promega and all other restriction enzymes from New England Biolabs.

Results

Sequence polymorphism in B163 parents at the *am* locus. The position of crossovers in the *am* region of Neurospora linkage group V has been determined in 84 prototrophs from a cross heteroallelic am^1 am^6, using both the conventional flanking

gene markers *sp* ≈6cM proximal and *his-1* ≈3cM distal and also close polymorphic restriction sites *HP* 8.3kb proximal and *HD* 5.7 kb distal (Cross B163 : Bowring and Catcheside, 1996 *ibid*). To obtain additional markers to map conversion tracts, the *am* coding region together with 311bp upstream and 582bp downstream (in total, 2381bp) were sequenced in the B163 parents using seven PCR primer pairs (Figure 11).

Of the nine natural polymorphisms, three are in the 311 bp upstream of *am*, one in intron I, a conservative base substitution in exon II, and three in intron II. No natural polymorphism was detected beyond intron II in the 1043bp of *am* coding sequences in exon III and only one in the 582bp of sequences downstream of the *am* stop codon.

Five of the nine sequence polymorphisms resulted in restriction site differences. These five sites, together with the am^1 and am^6 alleles provide seven readily detectable points of difference in the 1437bp from 302bp upstream to 1134bp into the *am* coding region.

Segregation of markers in B163 prototrophs. The location of markers is shown in Figure 11. The genetic constitution of the parents of cross B163, F6325 and F11089, together with the distribution of markers amongst their prototrophic progeny is detailed in Figure 14.

The composition of the local crossover group. As reported previously (Bowring and Catcheside, 1996 *ibid*) 14 B163 prototrophs had a crossover close enough to be considered associated with the conversion event at *am*. Eight of these 14 have a crossover point which appears separated from the converted region by a non-converted segment while in the remaining six there is an apparently uninterrupted transition from F6325 to F11089 DNA in *am* (Figure 14).

Relative proportion of am^6 and am^1 conversions. Those *am* prototrophs with the F6325 and F11089 association of *HP* and *HD* (the remote crossover and non-crossover groups, Figure 14) represent conversion of the 5' allele am^6 and the 3' allele am^1 respectively. Extrapolation of the data from the 27 of 145 sp^+ *his-1* progeny analyzed at the DNA level and correction for the fact that only 13 of the 14 *sp* $his\text{-}1^+$ and 28 of the 30 *sp his-1* prototrophs were available for such analysis (Bowring and Catcheside,

1996), suggests that 17.2 (9%) and 173.8 (91%) out of the total of 191 progeny that either experienced no crossover or a crossover remote from *am* were due to conversion of am^1 and am^6 respectively.

Polarity of conversion within *am*. The conversion frequency of each of the seven B163 alleles among am^6 convertants is shown in Figure 12. Data for progeny not analyzed at the molecular level were extrapolated as above. Conversion frequency peaks between am^6 and *Bd* and declines either side. 87% of am^6 convertants experienced co-conversion of *Bd*.

Conversion tract length. For am^6 conversions which form the larger class, markers either side of *am* make an estimate of minimum tract length possible. Each estimate is necessarily a minimum since exact positioning of tract termination points is not possible and many tracts end distal of *F* but proximal of *HD* which is 5.4kb away from *F*. The estimate must also be made with the caveat that selection for *am* prototrophs may select against longer tracts under some circumstances. For instance, an excision tract that covers the position of both mutant alleles on a given homologue will not yield a prototroph if initiation occurs to one side of the allele pair. The most frequent minimum length of conversion tracts among prototrophs is 741bp (Figure 13) although 60% of the tracts had a minimum length shorter than this and half of these had a maximum length shorter than 741 bp. By extrapolation, 12% of conversion tracts at *am* excluding those with an apparently associated crossover, were discontinuous.

Discussion

The segregation of sequence polymorphisms among prototrophs from cross B163 reveals information about the meiotic recombination event at *am*. The majority of events involved conversion of the am^6 allele and the conversion frequency of unselected alleles among such convertants suggests a peak of events immediately 5′ of this allele. Conversion tracts with a minimum length up to 741bp were detectable with the available markers. Although 40% were this length or more, at least 30% of tracts were shorter.

Smyth (Smyth D R (1970) Genetic control of recombination in the *amination-1*

region of *Neurospora crassa*, PhD Thesis, Australian National University) concluded, on the basis of data from more than 60 repulsion phase crosses of strains harboring various pairs of *am* alleles, that there is polarity across *am* with decreasing conversion frequency from 5′ to 3′. Our data on the relative conversion frequency of seven *am* alleles among am^6 convertants in cross B163 (Figure 12) confirm that conversion is most frequent at the 5′ end of the gene and in addition provides evidence that the incidence of conversion peaks 5′ of *am* and declines in frequency in both directions. A gradient of decreasing conversion frequency was also observed either side of the *ARG4* initiation site in *S. cerevisiae* (Nicolas A, Treco D, Schultes N P, Szostak J W (1989) An initiation site for meiotic gene conversion in the yeast *Saccharomyces cerevisiae* Nature 338: 35-39; Schultes N P, Szostak J W (1990) Decreasing gradients of gene conversion on both sides of the initiation site for meiotic recombination at the *ARG4* locus in yeast Genetics 126:813-822). The inferred conversion peak between am^6 and *Bd* suggests that conversion events are initiated in this interval, although the relative conversion frequency of these two alleles could also reflect variability in the position of initiation centered around *Bd*. The high frequency of am^6/*Bd* co-conversion (87%) is, however, consistent with initiation between these markers. Schultes and Szostak (1990) reported that markers either side of the *ARG4* initiation site co-converted at a frequency of between 64 and 91%.

A poorly repaired mismatch in *am*? In *S. cerevisiae*, the segregation of alleles that generate poorly repaired mismatches (PMS alleles) at *HIS4* (Detloff P, White M A, Petes T D (1992) Analysis of a gene conversion gradient at the *HIS4* locus in *Saccharomyces cerevisiae* Genetics 132: 113-123) and of usually well repaired mismatches in mismatch repair deficient strains at this locus (Reenan R A G, Kolodner R D (1992b) Characterization of insertion mutations in the *Saccharomyces cerevisiae* *MSH1* and *MSH2* genes: evidence for separate mitochondrial and nuclear functions Genetics 132: 975-985; Alani E, Reenan R A G, Kolodner R D (1994) Interaction between mismatch repair and genetic recombination in *Saccharomyces cerevisiae* Genetics 137: 19-39) and at *ARG4* (Alani et al., 1994) suggests an influence of

mismatch repair on gene conversion. In either case, polarity gradients are less steep. C/C pairings are among those mismatches that are poorly repaired in *S. cerevisiae* (Petes T D, Malone R E, Symington L S (1991) Recombination in yeast. In: Broach J R, Jones E W, Pringle J R (eds) The Molecular and Cellular Biology of the Yeast Saccharomyces, volume 1, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York). In cross B163, a C/C mismatch results from pairing of the non-transcribed strand of F11089 and the transcribed strand of F6325 at the *A* site. If, as is the case in *S. cerevisiae*, C/C mismatches are poorly repaired in Neurospora, this might explain two related observations on the segregation of *A* in B163 progeny. Firstly, where there is discontinuity in conversion tracts among am^6 convertants, the *A* site is always involved. Secondly, the conversion gradient from *Bp* to *A* appears shallower than either that from am^6 to *Bp* or that from *Bd* to *F* on the other, distal, side of the inferred conversion peak. *A* lies above the line describing the am^6 to *Bp* gradient but approaches it more closely when *A* conversions associated with discontinuous tracts are excluded (Figure 12).

The relationship between conversion and crossing over. The objective was to determine if there was anything unusual about conversion tracts at *am* that might explain the low level of association between conversion and crossing over at this locus. Tracts at *am* appear similar to those reported for *S. cerevisiae* and *S. pombe*. While some of the am^6 conversion tracts are at least 741bp long, it is possible that the average tract length exceeds this. As *am* prototrophs were selected for, longer tracts that extend proximal of am^6 and cover am^1, would not be recovered. Irrespective of this, the conversion tract lengths observed at *am* are not grossly different from those seen in *S. cerevisiae* (e.g. ~ 1.5kb, Borts R H, Haber J E (1987) Meiotic recombination in yeast: alteration by multiple heterozygosities Science 237: 1459-1465; Borts R H, Haber J E (1989) Length and distribution of meiotic gene conversion tracts and crossovers in *Saccharomyces cerevisiae* Genetics 123: 69-80) or for selected tracts among *S. pombe ADE6* prototrophs (~ 1kb, Grimm C, Bahler J, Kohli J(1994) *M26* recombinational hotspot and physical conversion tract analysis in the *ade6* gene of *Schizosaccharomyces pombe* Genetics 135: 41-51) where about half of events were reported to be associated with flanker exchange.

The demonstration that discontinuous conversion tracts are found not only in those recombination events initiated at *cog* but also in recombination events initiated by the hotspot 5' of *am* shows that any DNA juxtaposed to any recombination hotspot is likely to be diversified.

Methods and formulations for adaptation and use of the *his-3 cog* system of *Neurospora crassa* for *in vivo* diversification of heterologous DNA

Example 3 -- Culturing and Crossing Strains of *Neurospora crassa*

Strains of *N. crassa* can be cultured on a mineral medium, for example that of HJ Vogel (Am Naturalist 98: 435-446 1964), containing a carbon source (usually 2% sucrose), the vitamin biotin, and supplemented with any nutritional requirements imposed by the specific mutations present in the strains to be grown (DD Perkins *et al* Microbiol. Rev. 46:426-570 1982). Growth is at a temperature between 20°C and 36 °C , commonly 25 °C or 34 °C , depending on the presence of temperature sensitive mutants such as *cot-1* which confers colonial morphology at 34 °C but normal growth at 25 °C . The medium is usually solidified with agar such as Difco Bacto agar (2%) and dispensed into culture tubes closed with a gas permeable plug such as non-absorbent cotton. Following growth for 2 to 6 days, conidia, the asexual spores, are produced on the aerial hyphae (Figure 4). These can be collected with an inoculation loop and used to establish further genetically identical cultures. Aconidiate mutants are propagated by the transfer of mycelial fragments to fresh medium. *Neurospora crassa* has two mating types, *A* and *a,* determined by the idiomorphs *mat A* and *mat a* respectively (NL Glass and C Staben Fung. Genet. Newsl. 44:64, 1997). Crosses are made by co-inoculation of two strains, one of mating type *A* and the other of mating type *a* onto a medium formulated to have a low ratio of nitrogen to carbon source, such as that of M Westergaard and HK Mitchell (Am. J. Bot. 34:573-577 1947), SC medium (RH Davis and FJ de Serres Methods in Enzymol. 17A: 79-143 1970), or corn-meal medium. The crossing medium is supplemented with any specific nutritional requirements of the strains that result from the presence of specific mutations in one or both of the strains. Cross tubes are incubated at 25 °C as the sexual cycle is not completed

at higher temperatures in most laboratory strains of Neurospora as they contain a temperature sensitive tyrosinase. Crosses can also be made by inoculation of a single strain onto low nitrogen medium which is then incubated for 5 to 7 days at 25 °C to permit the development of protoperithecia by one parent (Figure 4). The protoperithecia are then fertilised by inoculating the culture with the second parent by dusting the mycelium with conidia or flooding with a suspension of conidia in water. The culture is then returned to 25 °C for incubation.

Fertilization, leading to the formation a dikaryon, and the subsequent events: karyogamy of pairs of nuclei of opposite mating type to form diploid cells, meiosis and the formation of octads of ascospores that are shot onto the wall of the culture tube, occur over a period of 14 to 40 days depending on the specific cross being made. A cross made in 4ml of medium in a 15x150mm pyrex tube normally yields about 5×10^7 ascospores, the products of ~6.10^6 separate meioses. Using the present invention, this is sufficient to yield between 10^2 and 10^5 variants of heterologous DNA depending upon the degree of difference between the two homologous sequences recombined, the recombinator employed and the selection method used to isolate recombinants.

Growing progeny from a cross are obtained as follows. Individual ascospore are isolated into tubes of appropriately supplemented Vogel's minimal medium containing 2% sucrose or another suitable carbon source, heat shocked at 60 °C for 30 minutes to break ascospore dormancy and kill any conidia present, and germinated at between 25 and 34°C. Alternatively, samples of ascospores are spread onto plates of appropriately supplemented Vogel's medium containing a carbon source designed to limit growth to compact colonies. This can be achieved with 1% sorbose and 0.1% sucrose or 1% sorbose and 0.05% glucose and 0.05% fructose. The spread ascospores are heat shocked and then incubated at 25 °C or, where both parents contain the *cot-1* mutation, incubated for 18 hours at 25 °C then at 34 °C until colonies are visible. The use of *cot-1* permits plating densities up to about 10/cm^2, which is about ten-fold higher than practicable when sorbose alone is used limit colony size. Individual progeny can be recovered from such plates by picking pieces of a colony aseptically onto an agar slope of appropriately supplemented Vogel's medium

containing 2% sucrose or another suitable carbon source and incubating, at 25 °C if *cot-1* is present, or up to 34 °C otherwise.

When one parent of a cross carries an auxotrophic mutation, ie a mutation requiring specific supplements to be added to the growth medium to compensate for their inability to synthesize the substance, half of the progeny of the cross will carry that mutation and will require suitably supplemented growth medium. If both parents carry different auxotrophic mutations, progeny having the mutant genes are recombined are the only ones that will grow in the absence of supplement added to the medium.

The frequency of prototrophic recombinants amongst the progeny of a cross can be estimated by preparing an ascospore suspension from the cross tube, plating aliquots of suitable dilutions of the suspension on both selective and fully supplemented medium and counting the colonies that grow following incubation of the plates at the appropriate temperature.

Example 4 -- Formulation of Plasmid Vectors

A plasmid can be constructed using methods common in the art for DNA manipulation, some of which are described above. One plasmid according to the invention includes the following DNA sequences in the following order:

1) The first included sequence is a sequence including the majority of the *his-3* gene, but lacking a portion of the sequence at the 5' end of the gene, typically a short portion of the sequence of about 30 to about 300 nucleotides, such that the start codon (starting at nucleotide 687 in Figure 7) is excluded from the plasmid. The first included sequence will terminate beyond the stop codon of the *his-3* gene (ending at position 3362 in Figure 7).

This first included DNA sequence can be derived either from a *his-3*$^{+}$ strain of *N. crassa* such as the St Lawrence 74A wild type (Figure 8), from the Lindegren wild type (Figure 7) which has a *his-3* gene that differs at 14 nucleotide positions from the sequence of St Lawrence 74A, or from one of a pair of complementing *his-3* mutants such as K26 derived from Lindegren and K458 derived from the Emerson *a* wild type that has a *his-3*

gene identical in sequence to that of St Lawrence 74A.

2) The second included sequence is optional. This sequence includes a promoter for the expression and control of expression of heterologous sequences that is functional in *Neurospora crassa*. This second included sequence is not required in certain circumstances, such as when this type of promoter is introduced as part of the heterologous DNA

3) The third included sequence is optional. This sequence, when the heterologous or foreign sequence codes for a messenger RNA, tags the messenger RNA transcribed from the heterologous DNA for export of the protein product. The third included sequence is present when such export is desired and when such a sequence is not already part of the heterologous DNA insert.

4) The fourth included sequence includes a cloning site having cleavage sites for one or more restriction enzymes. The presence of this sequence provides a site for insertion of heterologous genes and DNA sequences that are not present elsewhere in the plasmid.

5) The fifth included sequence is the cog^{L} allele of the *cog* recombinator. This sequence is typically from the 3' flank of the Lindegren wild type from nucleotide 5412 to 6831 (Figure 7) (preceding the start codon of *lpl*)

6) The sixth included sequence includes sequences within the *lpl* gene to provide homology downstream of *cog*. Such a sequence is typically from the sequence shown in Figure 7, from position 6831 and 3' for about several hundred base pairs.

7) The seventh included sequence provides a marker gene allowing either positive selection (for example hph^{R} which confers hygromycin resistance) or negative selection (for example mtr^{+} which confers p-fluorophenylalanine sensitivity) for the presence of the whole plasmid in Neurospora.

8) The eighth included sequence provides a selectable marker for the presence of the plasmid in *Escherichia coli*, for example amp^{R}.

9) The ninth included sequence provides a replication origin functional in *E. coli* to permit amplification of the plasmid in this species

The plasmid vector can also be constructed without one of more of the included

sequences. For example, simpler variants of this plasmid design include ones that omit *lpl* sequences and rely on cog^L sequences for correct register of the recombinational events (as explained hereinbelow) required to establish transplacement of chromosomal sequences with the desired construct which is: heterologous DNA flanked by an active *cog* allele and either $his\text{-}3^+$ or one of a complementing pair of *his-3* alleles such as K26 and K458.

Example 5 -- Cloning of Variants of Sequences to be Diversified

Two or more functional variants of the DNA to be diversified, each differing in sequence at multiple sites, are separately inserted into the multiple cloning site to form a panel of two or more plasmids that are identical except for the DNA incorporated into the multiple cloning site (Figure 9). The DNA can be inserted using methods known in the art. The foreign DNA variants can be derived from a variety of sources by methods known in the art. For example, the DNA variants can be derived from the genes of different species having homologous genes that produce proteins with equivalent function but differing properties such as pH tolerance, thermostability, substrate range, and the like, or, for example, from differently derived cell lines coding for monoclonal antibodies reacting to the same antigen, or incorporating sequences diversified by *in vitro* methods.

Example 6 -- Targeted insertion of heterologous DNA adjacent to cog^L

The Strains

The pair of strains used to derive the parents to be crossed together to effect *in vivo* diversification of heterologous sequences have the following characteristics:

1) The first characteristic is that one is of mating type *A* and the other mating type *a*, which allows mating.

2) The second characteristic is that both carry the same allele of each of the other ten known loci that determine heterokaryon compatibility, *het-c, -d, -e, -i,* and *het-5* to *het-10*. This allows the progeny of crosses to form heterokaryons in any combination of like mating type.

3) The third characteristic is that, when the heterokaryons are to be used to test some or all possible combinations of *in vivo* diversified DNA, each strain will carry

forcing markers for the heterokaryon, such as one or more auxotrophic mutations. This will be applicable where the heterologous DNA codes for a component of a protein having more than one subunit. If the protein is a heteromultimer, it will usually be appropriate to have the same forcing marker present in both parents and the two or more components of the heteromultimer will be diversified in separate crosses to yield the two or more panels of diversified sequences for use in combinatorial trials. If the protein is a homomultimer, there is the option of making two panels of diversified sequences in the same cross by making the parents heterozygous for the forcing markers. Suitable forcing markers include but are not limited to mutations that inactivate one of the following genes: *trp-2, pan-2, thi,* or *arg* leading to a requirement respectively for tryptophan, pantothenic acid, thiamine, or arginine.

4) The fourth characteristic is that the pair of strains can carry the mutation *tol* (DL Newmeyer Can. J. Genet. Cytol. 12:914-926 1970) which suppresses heterokaryon incompatibility between strains of different mating type to allow all combinations of progeny to form heterokaryons.

5) The fifth characteristic is that both strains carry an auxotrophic mutation of the *his-3* gene for selection of insertion of heterologous DNA between the *his-3* locus and cog^L. The allele chosen will be located towards the 3' end of the gene. Where selection for recombination at *his-3* will be used to enhance the yield of heterologous sequences diversified by recombination (see hereinbelow), the *his-3* alleles will not complement to ensure that any rare aneuploid progeny (progeny having two copies of the chromosome carrying the detectable gene) cannot give rise to a heterokaryon carrying both alleles that will no longer be auxotrophic for the trait of the detectable gene and thereby falsely mimic the desired recombinants. A suitable non complementing allele pair is K26 and K480. Suitable complementing pairs of *his-3* alleles are K26 and K458 or K480 and K458.

6) The sixth characteristic is that both carry cog^L and *lpl* sequences from the Lindegren strain to maximize homology with plasmid sequences and ensure that following insertion of heterologous DNA the strain has an active cog^L recombinator.

7) The seventh characteristic is that both carry *rec-2* to permit cog^L to cause recombination in *his-3* and the heterologous DNA.

8) The eighth characteristic is that both contain genes conferring resistance to any agent that is to be used to select against the presence of the whole plasmid. For example *mtr* if p-flurophenylalanine is to be used as the negative selective agent.

9) The ninth characteristic is that both will contain the mutation *cot-1* C102t where this is to be used to limit growth on plating media.

10) The tenth characteristic is that both contain such mutations and such additional sequences as can be required for optimum production and optimum secretion of the protein products of the heterologous genes or DNA sequences that are to be diversified.

Targeted Insertion

A plasmid construct carrying a different one of the two variants of the DNA sequence to be diversified is introduced into each member of a pair of parental strains by transfection of spheroplasts or electroporation of conidia by methods known in the art. The detectable genes in each parent strain are alleles that do not complement to ensure that any rare aneuploid progeny (progeny having two copies of the chromosome carrying the detectable gene) do not give rise to a heterokaryon carrying both alleles that will no longer be auxotrophic for the trait of the detectable gene and thereby falsely mimic the desired recombinants. *N. crassa* employed in the present invention preferably carry an *auxotrophic* mutant of the *his-3* gene. Typically, each allele chosen has a mutation toward the 3' end of the *his-3* gene. A suitable non complementing allele pair is K26 and K480. Suitable complementing pairs of *his-3* alleles are K26 and K458 or K480 and K458.

The plasmid is introduced so that the plasmid sequences replace those present in the chromosome (Figure 9 and Figure 10). The chromosome will have, in order: *his-3*$^+$ (or one of a complementing pair of *his-3* alleles such as K26 and K458 or K480 and K458), a variant of the DNA to be diversified, *cog*^{L} and the sequences proximal of *his-3* and distal of *cog*^{L} unchanged.

The transplacement can be achieved in one step by two reciprocal exchange events, one in the sequences between the 3' end of the *his-3* sequences in the plasmid and the location of the *his-3* mutation in the recipient strain (region 1) and one 3' of the

heterologous DNA contained in the plasmid and the 3' terminus of homology in the plasmid (region 2).

Alternatively, the transplacement can be achieved in two steps. First by an exchange in region 1, leading to insertion of the plasmid into the chromosome creating a duplication of those sequences shared by the plasmid and chromosome. This insertion can be selected for by plating the recipient cells on media selective for the positive selection marker (for example where the plasmid carries hph^R by the addition of hygromycin to the growth medium) as well as for the his^+ phenotype. Selection for the second event is effected by growing the cells on media without the positive selective agent, hygromycin where the plasmid carries hph^R, and screening for the absence of hph^R or by plating cells on media containing the negative selection agent, for example p-fluorophenylalanine if the plasmid carries the Neurospora mtr^+ gene which is dominant to *mtr* in vegetative cells carrying both alleles. Candidate strains are cultured individually in tubes of Vogel's mineral medium supplemented as required.

Correct transplacement of the sequences between the 5' end of *his-3* to distal of cog^L, including the heterologous sequences to be diversified is checked in each strain by the methods of southern transfer or PCR amplification and analysis of the position of sites sensitive to digestion by restriction enzymes or by PCR amplification and DNA sequencing.

Example 7 -- *In vivo* diversification of heterologous genes and DNA sequences

Parent variant 1 and Parent variant 2, being strains of different mating type, one *a* the other *A*, each carrying a different variant of the gene or DNA sequence to be diversified are co-inoculated into crossing tubes containing appropriately supplemented SC medium or another medium appropriate for the sexual stage in the life cycle. Where recombination at *his-3* is to be used to select for progeny in which the yield of recombinants is enhanced by ensuring that the whole of the heterologous DNA sequence was covered by a conversion tract, the *a* and the *A* strains will each carry a different *his-3* allele and the SC medium will contain histidine in addition to any other required supplement. The *his-3* alleles chosen will usually be a non-complementing pair, for

example one will be K26 and the other K480 (Figure 10), to ensure that rare heterokaryons arising by the breakdown of anuploids containing two copies of chromosome 1 that form when there is failure of chromosome disjunction in meiosis, will not mimic *his*$^+$ recombinants. If the alleles were a complementing pair such as K26 and K458 or K480 and K458, the resulting heterokaryon would grow on medium lacking histidine, mimicing a *his*$^+$ recombinant. Such complementing pairs are used to provide selection for transplacement of one *his-3* mutant with another in the process of inserting the heterologous DNA adjacent to *cog*. This preserves the presence of a *his-3* mutation when this is required for enhanced frequency of diversification of juxtaposed heterologous DNA.

Crossing tubes are incubated at 25 °C until ascospores are shot onto the wall of the tube and the ascospores matured by incubating for a further 7 days at 25 °C or for 48 hr. at 30 °C .

Strains in which the heterologous sequences have been diversified *in vivo* during meiosis are recovered by transferring individual ascospores to slopes of appropriately supplemented Vogel's N medium containing an appropriate carbon source such as 2% sucrose, heat shocking at 60 °C for 30 minutes to allow germination, and growing at 25 °C for 3 to 5 days. Where it is desired to enrich for diversified sequences by selecting only those progeny where a conversion tract extended from *cog*^{L} through the heterologous DNA into the *his-3* gene, ascospores are suspended in water, heat shocked at 60 °C for 30 min., spread on the surface of plates containing VM medium with 2% sorbose, 0.05% fructose and 0.05% glucose as carbon source and supplemented as is necessary for all auxotrophic mutants present excepting *his-3* (no histidine is added to the medium), and incubated for 18 hr at 25 °C followed by 24 hr at 34 °C , or where *cot-1* is absent, at 25 °C for 48 hr. Colonies are transferred individually to appropriately supplemented slopes of Vogel's minimal medium containing 2% sucrose or another suitable carbon source and grown at 25°C to yield a panel of strains enriched for *in vivo* diversified heterologous DNA sequences.

Example 8 – Use of Strains Containing *In Vivo* Diversified Heterologous DNA and Screening for New Variants Having Desirable Properties

Strains containing diversified heterologous sequences are used directly for expression of the variant gene or are combined in pairwise or higher order combinations in heterokaryons where a heteromeric protein such as an immunoglobulin is the product. Conditions and tests as needed for each specific case are used to screen for variants with the desired new combination of properties. Where *tol* is present in both parents of the cross used to diversify the heterologous sequences, all combinations of progeny can be combined in pairwise or in higher order combinations in heterokaryons. Where *tol* is absent from the parents, it is necessary to divide the progeny into two groups, those of mating type *a* and those of mating type *A*, and combine only those strains of like mating type in pairwise or higher order combinations.

The test for mating type can be effected by determining the mating reaction with tester strains one of *A* and one of *a* mating type RH Davis and FJ deSerres (Methods in Enzymol. 17A: 79-143 1970). In this method, strains of known mating type are inoculated onto SC medium and grown at 25 °C until protoperithecia are formed then inoculated with conidia of strains of unknown mating type. Mating type can be scored within 24-48 hours by the development of perithecia in tubes inoculated with conidia of the opposite mating type.

Example 9 – Provision of markers for the isolation of recombinant forms of heterologous DNA.

Targeting plasmid DNA into a specific chromosomal location is achieved by transfection of an auxotrophic mutant with a plasmid carrying non mutant sequence leading to the restoration of a normal gene in the recipient chromosome (Figure 9). Transformants with the plasmid DNA correctly targeted are selected by their ability to grow on media not supplemented with the requirements of the auxotrophic mutation. Where the fraction of progeny from a cross that experienced conversion of heterologous DNA is too low to provide a good yield of diversified sequences, it is necessary to be able

to enrich for those that did experience conversion to provide the panel of diversified sequences. This can be achieved if a *his-3* mutation remains in the chromosome of the recipient cell after insertion of the heterologous DNA.

Complementing allelic mutations of the *his-3* gene (DG Catcheside and T Angel Aust. J. Biol. Sci. 27:219-29 1974) provide a means of achieving transplacement at *his-3* leaving a *his-3* mutation conferring a requirement in the chromosome carrying the heterologous DNA. This is made possible because *his-3* codes for a peptide that forms a homomultimeric protein. Although the homomultimer formed of subunits coded by *his-3* K26 and the homomultimer formed by subunits coded by *his-3* K458 are both inactive, a heteromultimer containing subunits coded by *his-3* K26 and other subunits coded by *his-3* K458 is enzymatically active. As a result, a heterokaryon containing nuclei having the *his-3* K26 mutation and also nuclei having the *his-3* K458 mutation is able to grow on media that do not contain histidine. Since K26 is located 3' of K458 and since Neurospora cells are usually multinucleate, transfection of strain carrying the *his-3* K458 mutation (and having a mating type and such other genetic markers as can be required for screening of the DNA following diversification) with a targeting plasmid carrying heterologous DNA and the *his-3* K26 sequence, instead of the *his-3* wild type sequence, will establish a heterokaryon with two sorts of nuclei. One nuclear type will carry the *his-3* K458 mutation and the other type will carry the heterologous DNA juxtaposed to the *his-3* K26 mutation. The heterokaryon can be specifically selected by its ability to grow on media that contain no added histidine.

A homokaryon containing only the nuclear type carrying the *his-3* K26 mutation and the juxtaposed heterologous DNA is isolated by picking colonies to slopes of minimal medium, growing to conidia and isolating homokaryons by establishing new cultures on medium containing histidine from single conidia. Random assortment of the two nuclear types into the conidia frequently gives rise to homokaryons. This can be made more efficient by forcing the production of microconidia having only a single nucleus by growing the heterokaryon on media supplemented with iodoacetate (D Ebbole and MS Sachs Fun. Genet. Newsl. 37:17-18 1990).

The homokaryon can then be used as one parent [parent (variant 1)] of the cross

used to diversify the heterologous DNA (Figure 10). A second homokaryon containing the *his-3* K480 mutation and foreign variant 2 of the heterologous DNA can be isolated in a similar way (Figure 10) by transfecting a *his-3* K458 strain, of opposite mating type to that used to construct parent (variant 1) (and with such other genetic markers as can be needed for subsequent screening of the DNA following diversification). The plasmid used for transfection will carry a homolog of the DNA to be diversified, different to that used for construction of parent (variant 1), juxtaposed to the DNA sequences from the *his-3* gene from a *his-3* K480 strain. Since *his-3* K480 is located 3' of *his-3* K458 and they are a complementing pair, a heterokaryon with two sorts of nuclei can be selected on minimal medium. One nuclear type will carry the *his-3* K458 mutation and the other type will carry the variant 2 DNA juxtaposed to the *his-3* K480 mutation. Parent (variant 2) for the DNA diversification cross which carries variant 2 DNA juxtaposed to *his-3* K480 can be recovered from the heterokaryon in the same manner as that used to establish parent (variant 1).

The *his-3* alleles used for the construction of parent (variant 1) and parent (variant 2) are dictated by (1) the plasmid sequence should carry the most 3' *his-3* allele (2) the pair used for the establishment of the heterokaryon must be a complementing pair and (3) the pair used for parent (variant 1) and parent (variant 2) should be non complementing. K26 and K480 do not complement. Dictate (1) to ensure *his-3*$^+$ chromosomes are not established, dictate (2) to provide for the selection of chromosomes carrying a *his-3* mutant juxtaposed to the heterologous DNA to be diversified and dictate (3) to eliminate the selection of *his-3*$^+$ recombinants from the diversification cross which are aneuploid. (aberrant cells carrying two copies of the section of chromosome bearing the *his-3* gene).

The invention has been described with reference to various specific and preferred embodiments and techniques. However, it should be understood that many variations and modifications can be made while remaining within the spirit and scope of the invention.

All publications and patent applications in this specification are indicative of the level of ordinary skill in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each

individual publication or patent application was specifically and individually indicated by reference.